基于叠层循环神经网络的语义关系分类模型*

郝志峰[1,2]，陈培辉[1]，蔡瑞初[1]，温　雯[1]，王丽娟[1]

(1. 广东工业大学 计算机学院，广州 510006; 2. 佛山科学技术学院 数学与大数据学院，广东 佛山 528000)

摘　要：基于循环神经网络结合句法结构的方法被广泛运用于关系分类，利用神经网络对输入的编码信息自动获取特征并实现关系分类；然而，目前已有的方法主要是基于单一特定句法结构的模型，而特定句法结构的模型不能够迁移到其他句法结构类型上。针对该问题，提出一种融合多句法结构的叠层循环神经网络模型。该叠层循环神经网络分为两层进行网络构建，首先在序列层进行实体预训练，通过 Bi-LSTM-CRF 融合 Attention 机制，提高模型对文本序列上实体信息的关注度，从而获取更加准确的实体特征信息，促进关系层阶段更好地分类；其次在关系层，将 Bi-Tree-LSTM 嵌套在序列层之上，并将序列层的隐状态与实体特征信息传入关系层，利用共享参数对三种不同的句法结构进行加权学习，通过端到端的模型训练并实现语义关系分类。实验结果表明，该模型在 SemEval-2010 Task8 语料库上的 marco-F1 值达到了 85.9%，并进一步地提升了模型的鲁棒性。

关键词：叠层循环神经网络；多句法结构；Bi-Tree-LSTM；注意力机制；关系分类

中图分类号：TP183　　doi: 10.19734/j.issn.1001-3695.2018.06.0461

Sematic relation classification model via hierarchical recurrent neural network

Hao Zhifeng[1, 2], Chen Peihui[1], Cai Ruichu[1], Wen Wen[1], Wang Lijuan[1]

(1. *Faculty of Computer*, *Guandong*, *University of Technology*, *Guangzhou* 510006, *China*; 2. *School of Mathematics & Big Data*, *Foshan University*, *Foshan Guangdong* 528000, *China*)

Abstract: The method based on recurrent neural network combined with syntactic structure is widely used in relation classification, and the neural network is used to automatically acquire features and realize relation classification. However, the existing methods are mainly based on a single specific syntactic structure model, and the model of a specific syntactic structure cannot be transferred to other types of syntactic structures. Aiming at this problem, a hierarchical recurrent neural network model with multi-syntactic structure is proposed. The hierarchical recurrent neural network is divided into two layers for network construction. Firstly, entity pre-training is performed in the sequence layer. The Bi-LSTM-CRF fusion Attention mechanism is used to improve the model's attention to the entity information on the text sequence, thereby obtaining more accurate. The more accurate entity feature information promotes better classification in the relation layer stage. Secondly, in the relation layer, the Bi-Tree-LSTM is nested above the sequence layer, and the hidden state and entity feature information of the sequence layer is passed into the relation layer, then three different syntax structures are weighted learned using the shared parameters and classify the semantic relation finally. The experimental results show that the model has a marco-F1 value of 85.9% on the SemEval-2010 Task8 corpus, and further improves the robustness of the model.

Key words: H-RNN; multi-syntactic structure; Bi-Tree-LSTM; attention; relation classification

0　引言

为了应对互联网产生的海量的非结构化文本数据，通常将其转换为结构化的数据，从而帮助人们更快速地获取其中重要的信息。关系分类[1]作为其中使用最广泛的技术之一，越来越多的受到了研究界的关注。而关系分类主要涉及到文本序列和句法依赖两类重要的信息，如何引入结构化知识，将两者进行更好的融合，去提高分类效果，是本文的研究重点。

收稿日期：2018-06-20；**修回日期：**2018-08-10　**基金项目：**NSFC-广东联合基金资助项目（U1501254）；广东省自然科学基金资助（2014A030306004，2014A030308008）；广东省科技计划项目（2015B010108006，2015B010131015）；广东特支计划资助项目（2015TQ01X140）；广州市珠江科技新星资助项目（201610010101）；广州市科技计划项目（201604016075）

作者简介：郝志峰（1968-），男，江苏苏州人，教授，主要研究方向为机器学习、人工智能（mazfhao@scut.edu.cn）；陈培辉，男，硕士研究生，主要研究方向为自然语言处理、机器学习；蔡瑞初，男，教授，主要研究方向为因果关系、机器学习、数据挖掘；温雯，女，副教授，主要研究方向为文本挖掘、机器学习；王丽娟，女，副教授，主要研究方向为高维数据聚类分析.

近年来，随着深度学习的高速发展，关系分类的做法正在由传统的基于构造特征工程[2-4]和构造核函数[5-7]的方法，逐渐过渡到直接使用深层神经网络对输入的编码信息进行学习并实现分类。其中，代表性的工作包括基于循环神经网络(recurrent/recursive neural network)和卷积神经网络(convolutional neural network)的方法，而基于 RNN 的模型在结构上更适应文本问题，它能够直接地表示语言结构，如文本序列和句法依赖树结构。

在关系分类过程中，主要包括两阶段的任务，分别是命名实体识别和语义关系分类，对应的主流方法分别是基于序列标注[8-10]的方法和基于句法结构[11-13]的方法。在进行第二阶段的任务时，目前主要基于单一特定的句法结构，这类模型存在一定的局限性，只能够处理特定的句法结构，不能够很好地迁移到其他句法结构上。比如，对于序列“A thief who tried to steal the truck broke the ignition with screwdriver.”，其最短路径结构(SDP)包含了实体对“thief”“ignition”以及实体对之间的谓语成分“broke”；对于同样的序列，其子树结构(SubTree)不仅包含了 SDP，同时包含了其他成分信息如定语“A”，修饰成分“with screwdriver”等。因此，SDP 的网络模型对于 SubTree 结构会存在信息丢失的问题，模型上不能够直接进行迁移。

因此，本文提出一种融合多句法结构的叠层循环神经网络模型。a)在序列层进行实体预训练，利用 Bi-LSTM-CRF 融合 attention 机制提高模型对实体信息的关注度，更加准确地获取实体标签信息；b)在关系层，利用 Bi-Tree-LSTM 接收序列层输出和依赖标签的融合特征作为新的输入，对三种不同的句法结构进行加权学习，利用后向传播对共享参数进行更新，最终通过 softmax 分类器输出语义关系类别。考虑到两阶段任务之间的交互是密切的，通过构造端到端的模型，能够互相促进和提升完整模型的效果。

本文研究的意义在于，提出了一种融合多句法结构的网络模型，并且对两个任务构造叠层的端到端网络，能够更好的适应不同的句法结构，提高了关系分类的精度和稳定性。针对不同层次的网络，其贡献在于：a）在序列层的实体预训练阶段，考虑了输入与输出间的相关性，结合 word-level attention 机制能够有效提升模型对文本序列上实体信息的关注度，这对于关系分类起到了促进作用；b）在关系层时，将多种句法结构进行融合，使得模型能够在同一个框架中处理不同的句法结构进行有效的分类，提升了模型的鲁棒性；同时，验证了不同的句法结构对于关系分类的贡献度。

1 研究现状

关系分类本质上是一个分类问题，通常的做法是先识别出一个句子中存在的实体对，然后利用分类器决定哪些成分是真正需要的关系。早期的关系分类问题通过借助知识库[14]来解决，由于构建知识库的代价成本过大，因此将研究方向转向了机器学习。当前的研究可以分为如下三类：

a)基于特征的方法。通过抽取大量语言学特征，包括语义和语法，将其进行组合形成特征向量集，利用分类器（如最大熵模型和 SVM）进行分类[2-4]。该方法在处理特定领域时效果较好，但在特征集的选择和设计上依赖于专家知识，需要花费的时间成本较大。

b)基于核的方法。该方法通过计算两个对象在高维稀疏空间上的内积以获取结构化特征。Zelenko 等人[5]通过设计树核函数进行浅层的句法分析以获取结构共性；Culotta 等人[6]扩展了 Zelenko 的工作，将树核函数扩展到了依存关系树上，并结合了语法分析信息；Bunescu 等人[7]将句法最短路径与核函数进行融合，探索在不同句法结构上的分类效果；Zhang 等人[15]利用卷积树核探索句法特征在关系分类上的作用；Zhou 等人[16]在文献[15]的基础上添加了文本内容信息。该类方法在核函数的选择上需要技巧性，大数据量时训练速度慢；同时分类性能依赖于 NLP 工具，而文本预处理的错误结果会影响分类器性能。

c)基于神经网络的方法。该方法的优势在于对输入的编码信息进行自我学习，无须手动构建特征，同时丰富的编码信息克服了传统方法的稀疏问题。在 RNN 结合句法结构的方法上，Socher 等人[17]基于递归矩阵-矢量的方式获取语句在组成角度上的语义；Xu 等人[11,18]和 Liu 等人[19]证明了句法最短路径对于神经网络模型获取语义关系是有帮助的；Li 等人[20]讨论了不同的句法树结构在神经网络模型中的分类效果；Miwa 等人[21]提出了 Bi-Tree-LSTM，并同时考虑了不同句法结构类型及其孩子节点的数量关系；Zhou 等人[12]直接在 Bi-LSTM 上嵌套了一层 attention，增强了编码后的权重信息；Xiao 等人[22]将长句子进行切割，并用两层的 RNN 网络结合 attention 对不同层的信息进行编码；Zhang 等人[23]提出 RNN 结合 CNN 的改进方法，将 CNN 接在双向 LSTM 上，对经过 attention 的权重卷积后进行全连接输入分类器。而随着增强学习和对抗网络的提出，也有一些相关的研究成果，Feng 等人[24]提出了一种利用增强学习处理噪声数据的关系分类方法，Liu 等人[25]提出了一种关系分类的对抗训练框架。

相比较其他两类传统方法，基于神经网络的方法优势在于网络能自主学习获取特征，无须人工定义特征，且分类效果目前达到最优。相对于 CNN，基于 RNN 网络的模型，能够更好的处理文本序列和句法结构。但当前结合句法结构进行关系分类的模型，存在句法结构单一的问题，无法将特定句法结构的模型迁移到其他不同的句法结构上。因此，本文提出了一种融合多句法结构的叠层循环神经网络模型，分别利用 Bi-LSTM-CRF(融合 Attention 机制)和 Bi-Tree-LSTM 对文本序列和多种不同的句法结构进行学习，通过共享参数进行端到端的训练，最终输出语义关系类别。

2 基于 H-RNN 的关系分类框架

本文的关系分类框架主要包括四部分，分别是输入序列、实体预训练、多句法融合的结构和输出语义关系。框架主体部

分在于序列层和关系层上，基于双向循环神经网络来表示文本序列与句法结构，并将关系层嵌套于序列层之上，构成端到端的模型，模型框架如图 1 所示。更具体的是，在序列层，输入的文本序列经过 Bi-LSTM 进行编码，通过 word-level 的 Attention 获取每个字在句子中的权重，最后利用 CRF 对其进行双向解码；在关系层，将依赖标签与序列层的输出进行拼接，作为此阶段的输入，经过 Bi-Tree-LSTM 对多种句法结构进行加权学习，获取实体对的候选关系，最后通过 softmax 分类器获取最终的语义关系。

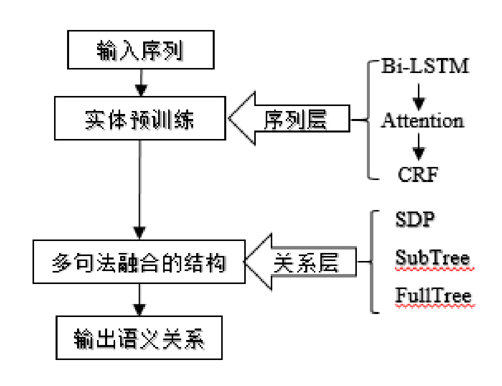


图 1 基于 H-RNN 的关系分类框架

Fig.1 H-RNN based relation classification framework

2.1 框架的基本阐述

本文提出的模型框架主要由三部分构成，分别是输入序列的词向量表示、序列层的实体预训练、关系层的多句法融合结构。

词向量包含维度为 d_w、d_p、d_d 的字 v_w、词性标签 v_p、句法标签 v_d，均为经过预训练的 embedding。

在序列层部分，本文采用标准的 Bi-LSTM 对序列进行编码[26]。在 t 时刻，LSTM 包含了一个输入门 i_t、一个忘记门 f_t、一个输出门 o_t、一个记忆单元 c_t 以及一个隐状态 h_{t-1}，其计算公式如下：

$$
\begin{aligned}
i_t &= \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}) \\
f_t &= (W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}) \\
o_t &= \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}) \\
u_t &= \tanh(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}) \\
c_t &= i_t \odot u_t + f_t \odot c_{t-1} \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
\tag{1}
$$

其中：σ 是逐元素的 sigmoid 函数，$\odot$ 是逐元素的积，W 和 U 是权重矩阵，b 是偏置向量。

对于一个包含 n 个字的序列 $(x_1, x_2, \cdots, x_n)$，每个字都用 d 维的特征向量来表示。在 t 时刻，对于当前的第 t 个字，LSTM 接收当前的输入向量为 $x_t = [v_w, v_p]$，前一刻的隐状态 h_{t-1} 和细胞状态 c_{t-1}，并返回当前的隐状态 h_t。由于在这里使用的是 Bi-LSTM，因此对于每个字在前向和后向均获得了一个隐状态，分别是 $\overrightarrow{h_t}$ 和 $\overleftarrow{h_t}$，经过拼接得到 $s_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$，作为传入关系层的输入之一。

在序列层的解码阶段，采用标准的 CRF 模型对来自 Attention 的输出向量进行解码，标记模式为 BILOU[27]，其中每个实体标签包含了实体类型和实体位置信息。如图 2 所示，B-PER 和 L-PER 表示 Sidney Yates 是一个 PER 实体类型以及相应的位置。最终得到的实体标签向量 v_e 将作为关系层的输入之一。

关于模型在序列层的 Attention 部分以及关系层将会在下文分开具体阐述。

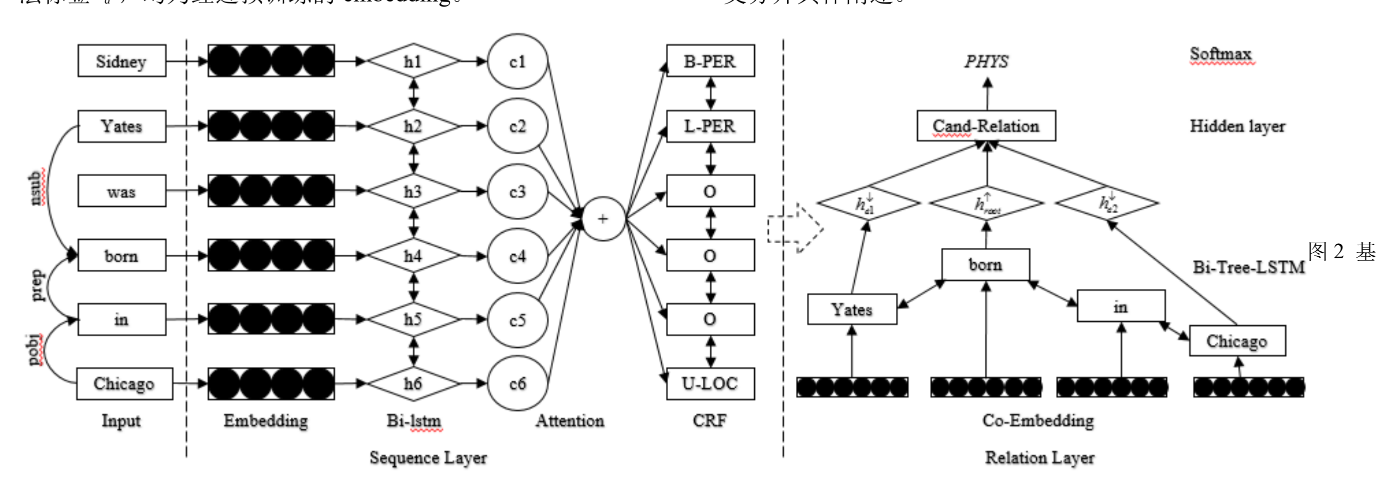


图 2 基于叠层循环神经网络的关系分类模型(当模型选择 SDP 结构时)

Fig.2 Semantic relation classification model via hierarchical recurrent neural network(when model choose SDP)

2.2 Word-level Attention

序列层不仅蕴涵了句子上的实体信息，同时包含了实体的上下文信息以及部分非相关信息。为了使模型能够更好的预测实体间的关系，需要让模型在这一层上更多的聚焦于序列上最关键的信息，也就是实体与谓语动作这两类主体信息。

因此，模型在序列层部分引入了一种字级别(word-level)的 attention 机制。Attention 机制使得模型能够沿着文本序列逐字地处理 Bi-LSTM 所产生的隐藏向量 s_t，从而获取到相应的权重分配，产生它们的加权表示 r，如式(2)所示。

$$
\begin{aligned}
z_t &= \tanh(W^{(w)}h_t) \\
\alpha_t &= \frac{\exp(v_z^{\mathrm{T}} z_t)}{\sum_{j=1}^{T}\exp(v_z^{\mathrm{T}} z_j)} \\
r &= \sum_{t=1}^{T} a_t h_t
\end{aligned}
\tag{2}
$$

其中：$W^{(w)}$ 是权重矩阵，v_z 是权重向量，v_z^{T} 是该权重向量的转

置。

2.3 关系层

关系层表示句法依赖树上一对实体对之间的候选关系，如图 1 的右部分所示。已有的实践证明，不同的句法结构对于关系分类有不同程度的效益，但目前的模型基本都是基于特定句法结构下的研究。因此，本文提出将三种不同的句法结构进行加权融为一体，让网络学习在不同的句法结构下对关系进行分类。

首先，明确 SDP、SubTree、FullTree 结构的定义：

a)SDP,最短路径是指最近公共节点和两个实体目标词之间的核心依赖路径；

b)SubTree,子树是包含了最短路径以及最近公共节点下的子树的句法结构；

c)FullTree,全树指完整的句法依赖树，能够捕获到完整句子的上下文信息。

本文采用 Bi-Tree-LSTM 获取实体对之间的候选关系，该网络结构不仅能够充分学习底部叶子节点及其孩子节点的信息，同时还能够将顶部的根节点信息传递到底部叶子节点，这对于网络充分学习句法结构上的信息是有益的。Miwa 等[21]的模型通过将句法结构定义为两种类型，并通过一个映射函数来处理两种类型的句法结构树。基于此，本文提出一种改进的 Bi-Tree-LSTM，通过共享权重融合三种句法结构信息，并能够计算不同结构下节点的孩子节点数量。

$$
\begin{aligned}
i_t &= \sigma\left(W^{(i)}x_t + \sum_{n\in C(t)}\sum_{s\in T} p_s U_s^{(i)} h_{tn} + b^{(i)}\right)\\
f_t &= \sigma\left(W^{(f)}x_t + \sum_{n\in C(t)}\sum_{s\in T} p_s U_s^{(f)} h_{tn} + b^{(f)}\right)\\
o_t &= \sigma\left(W^{(o)}x_t + \sum_{n\in C(t)}\sum_{s\in T} p_s U_s^{(o)} h_{tn} + b^{(o)}\right)\\
u_t &= \tanh\left(W^{(u)}x_t + \sum_{n\in C(t)}\sum_{s\in T} p_s U_s^{(u)} h_{tn} + b^{(u)}\right)\\
c_t &= i_t \odot u_t + f \odot c_{t-1}\\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
\tag{3}
$$

其中: $\sum_{n\in C(t)}\sum_{s\in T} p_s U_s^{(o)} h_{tn}$ 表示对句法结构进行线性加权的学习，当模型学习到不同的句法结构时，需要计算该字 x_t 在相应的句法结构下的孩子节点数量， $s\in T$ 表示结构 s 为所有树结构 T 下的一种。本文定义两种节点类型，一种是 SDP 路径上的节点，另一种是 SubTree 和 FullTree 上除了最短路径之外的其他所有节点。

由于实体识别与语义关系分类是联合任务，通过将关系层嵌套在序列层之上，将文本序列获取的有效信息传递到关系层上，进行端到端的训练，能够有效的提升完整模型的训练效果。

关系层上的 LSTM 单元在第 t 个字接收的输入为 $x_t = \left[s_t, v_t^{(e)}, v_t^{(d)}\right]$ (即图 2 右部分中的 Co-Embedding)，分别为该字在序列层的隐状态向量 s_t 、实体预训练获得的实体标签 $v_t^{(e)}$ 、该字的依赖标签 $v_t^{(d)}$ 。

在完整模型的解码阶段，基于组合的思想对每个检测到的实体的最后一个字进行组合以获取最终的候选关系，即在 BILOU 模式下标记为 L 和 U 标签的字。在图 1 示例中，对 Yates(L-PER)和 Chicago(U-LOC)标记一个候选关系。对于每一个输入的句子，关系层经过 Bi-Tree-LSTM 获取候选关系 d_p，经过包含 tanh 函数的隐含层，最终通过 softmax 层预测关系标签。

关系层使用的是 Bi-Tree-LSTM，因此候选关系 $d_p = \left[h_{root}^{\uparrow}, h_{e1}^{\downarrow}, h_{e2}^{\downarrow}\right]$，其中，“ $h_{root}^{\uparrow}$ ”代表自底向上方向的顶部单元的隐含层状态向量（即，谓语成分），“ $h_{e1}^{\downarrow}$， $h_{e2}^{\downarrow}$ ”代表自顶向下方向在两个底部叶子节点的隐含层状态向量（即，实体节点）。在最终的关系预测阶段，模型通过两层的神经网络预测最终的候选关系，分别是一个隐含层 $h_p^{(d)}$ 和 softmax 层，公式如下：

$$
\begin{aligned}
h_p^{(d)} &= \tanh\left(W^{(d_h)} d_p + b^{(d_h)}\right)\\
P &= \mathrm{softmax}\left(W^{(d_y)} d_p + b^{(d_y)}\right)
\end{aligned}
\tag{4}
$$

2.4 模型训练

对于模型的训练过程和参数设置问题。在序列层上，采用标准的线性链 CRF，基于 BILOU 模式对训练集进行实体预训练，将得到的标签作为关系层的输入之一。在关系层上，本文将 softmax 作为分类器，置于关系层的隐含层之上，接收来自隐含层的句子表示 $h_p^{(d)}$ 并产生该句子的语义关系概率分布 P (如式(4)所示)，其训练目标是最小化介于句子的预测关系与实际语义关系之间的交叉熵，即：

$$
loss = -\sum_{k=1}^{n}\sum_{i=1}^{C} y_{ki} log\left(y_{ki}^{'}\right) + \frac{\lambda}{2}\|\theta\|^2
$$

其中: y_{ki} 表示句子 k 属于类别 i 的概率， $y_{ki}^{'}$ 表示句子 k 预测为类别 i 的概率， λ 是 L2 正则化项， θ 是它的参数，本文只对权重 W 和 U 进行正则化。

同时为了防止过拟合，设置了训练提前结束，具体是设定一个数值 n=100，交叉熵在 n 次迭代仍没有达到之前的最好值就可以提前结束训练。为了保证网络训练过程的一致性，对长度不等的句子进行填充。模型训练过程详细的参数设置在 3.2 节中阐述。

3 实验结果与讨论

实验旨在证明：a)融合 Word-level 的 Attention 机制能够有效提升模型对于文本序列上实体信息的关注度，对完整模型的分类效果起到促进作用；b)多种句法结构的加权学习能提升模型的鲁棒性，保证模型能够处理不同的句法结构。下面，首先介绍实验数据与评价指标，实验设置，然后分别测试实验和结果分析，最后与其他方法进行对比。

3.1 实验数据与评价指标

实验采用 Semeval-2010 Task 8[1]作为实验数据集，该数据集是评判关系分类任务的经典数据集，共包含 8000 条训练样

本和 2717 条测试样本，关系类别包括 9 种语义关系以及 Other 关系（即实体对之间不存在关系），具体语义关系-样本分布如表 1 所示。

表 1 SemEval 2010 Task8 语义关系-样本分布

Table 1 Semantic relation-sample distribution of SemEval 2010 Task8

关系类型	训练集	测试集
Other	1410	454
Cause-Effect	1003	328
Component-Whole	941	312
Entity-Destination	845	292
Entity-Origin	716	258
Product-Producer	71	231
Member-Collection	690	233
Message-Topic	634	261
Content-Container	540	192
Instrument-Agency	504	156
合计	8000	2717

实验采用 marco-F1 作为评价指标，该计算依赖于正确率(precision)和召回率(recall)，计算公式如下：

$$
\begin{aligned}
&Precision = \frac{True\ Positive + True\ Negative}{Positive + Negative} \\
&Recall = \frac{True\ Positive}{Positive} \\
&F1 = \frac{2 * precision * recall}{precision + recall}
\end{aligned} \tag{5}
$$

其中: $True\ Positive + True\ Negative$ 表示输出的判断正确的关系个数；$Positive + Negative$ 表示输出的所有的关系个数；$True\ Positive$ 表示在输出判断正确的关系中，属于 $Positive$ 类别中的关系个数，$Positive$ 表示所有关系中正确类别的关系个数。

3.2 实验设置

词向量采用由 Glove2vec 对英文维基百科预训练的词向量进行初始化，维度为 200；采用 Stanford Dependency Parser 对句子进行句法依赖解析。模型的激活函数选用 tanh 函数，隐含层节点数为 200，采用 softmax 作为本文的分类器，对权重 W 和 U 采取 L2 正则化。为防止过拟合，在训练过程中分别对序列层和关系层引入 dropout 策略，dropout 值取 0.3。另外，采用批量的 Adam 和梯度剪裁用于模型训练更新权重参数，批量大小取 1，训练轮数取 100。对于其他参数，均采用随机初始化处理。

3.3 实验结果与分析

本文实验以文献[21]中的结果为基准，该方法在未添加额外的先验知识 WordNet 时，能够达到的分类效果为 84.4%。

3.3.1 序列层的改进

本文在第一组实验中，基于文献[21]的方法，通过改进该模型的序列层，分别是使用了不同的解码方式以及添加 word-level 的 Attention 机制，得到实验结果如表 2 所示。

由实验结果可知，将解码方式改进为 CRF 并添加 Attention 机制，提高了基准模型最终的分类效果。其原因在于，在原先的基准模型的序列层上，实体预训练的解码方式是基于单向的贪心解码，通过将其替换成 CRF，能够捕获双向的上下文信息，从而学习到更充分的信息。另一方面，尽管替换为 CRF 解码方式，但模型的分类效果并没有显著提升，分析原因可能是对于实体信息的学习不够充分。通过添加 word-level 的 attention 机制提高了模型的分类效果，其原因在于提高了模型对于文本序列上实体信息的关注度，让模型更好的学习到实体信息。该实验说明，使用 CRF 和 attention 机制能够有效提升模型在实体预训练的效果。

表 2 改进序列层的实验结果(未添加额外的先验知识)

Table 2 Experimental results of improved sequence layer(without extra prior knowledge)

方法	Marco-F1(%)
Baseline（SDP）	84.4
+CRF	84.5
+CRF, +ATTENTION	**84.9**

3.3.2 融合多句法结构

在第二组实验中，保持文献[21]进行实体预训练的序列层不变，通过将基准模型的依赖层替换成融合多句法结构的关系层，实验结果如表 3 所示。

表 3 融合多句法结构的实验结果（添加了额外的先验知识，WordNet）

Table 3 Experimental results of multi-syntactic structure(with extra prior knowledge, WordNet)

	Baseline	多句法结构
SDP	85.1	85.3
SubTree	83.9	84.0
FullTree	82.9	83.0

基准模型的句法结构采用映射函数来选择使用何种句法结构，而本文的模型通过融合多种句法结构为一体，其目的在于提升模型的鲁棒性，使得模型能够直接处理不同的句法结构。由实验结果可知，本文提出的融合多句法结构的模型，对于不同的句法结构都有着稳定可靠的表现，其结果较基准模型有所提升，具备良好的鲁棒性。

进一步地，由第一组实验已知，通过使用 CRF 解码并添加 Attention 机制的方式，能够提升序列层的实体训练效果，这对于模型进一步分类起到了促进作用。因此，通过将两组实验进行结合，即将改进的序列层与多句法融合的结构叠层嵌套，对完整模型进行端到端的训练，实验结果如表 4 所示。

在第三组实验中，本文提出的 H-RNN 模型要优于添加了额外先验知识的基准模型和基于序列的模型。其原因可能是：a) 实体识别与关系分类是互相关联的任务，序列层的改进，提升了实体预训练的效果，对于训练端到端的完整模型起到了促进作用，从而提升了分类效果；b）文本序列和句法结构都蕴含了不同的信息成分，基于序列的模型在信息量上过于单一，通过将结构化的信息引入到文本序列上，能够丰富模型的分类能

力；c）同时，进一步证明了，SDP 在序列的关系分类问题上优于 SubTree 和 FullTree。因为 SDP 在结构上要更直观简洁，其他两种结构部分多余的节点信息会引入一些冗余，可能会导致训练过程对重要信息的学习造成干扰。

表 4　改进的序列层+融合多句法结构的实验结果（添加了额外的先验知识，WordNet）

Table 4　Experimental results of improved sequence layer & multi-syntactic structure(with extra prior knowledge, WordNet)

	多句法结构
SDP	**85.9**
SubTree	84.7
FullTree	83.6
SDP(baseline)	85.1
SPSeq	84.4
SPXu	84.7

表 5　H-RNN 与其他方法进行对比

Table 5　Comparison between H-RNN with other methods

模型	添加的特征	Marco-F1/%
SVM	POS, WordNet, prefixes and other morphological features, dependency parse, Levin classes, PropBank, FanmeNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
CNN	position features, words around nominals, WordNet	82.7
SDP-LSTM	POS embeddings, WordNet embeddings, grammar relation embeddings	83.7
BLSTM	POS, NER, WordNet, position features, dependency feature, relative-dependency feature	84.3
ATT-BLSTM	Without any other features, only use word vector(50 dim & 100 dim) and position indicators	84.0
2ATT-BLSTM-BLTM	position features, WordNet, NER	84.3
BLSTM-BTLSTM(SDP)	POS, dependency feature, WordNet	85.1
H-RNN(SDP)	POS, dependency feature, WordNet	**85.9**

3.4　与其他方法对比

本文选取了当前在关系分类中最好的一些方法与本文进行横向对比，分别是：

a)SVM[28]。利用大量人工定义和工具抽取的特征，使用 SVM 进行训练和分类。

b)CNN[13]。将句子作为序列数据处理，并利用卷积神经网络学习句子级别的特征；同时，利用一个特殊的位置向量来表示序列上的每一个字。通过将句子级别特征和词法特征拼接在一起，输入分类器 softmax 中进行预测。

c)SDP-LSTM[11]。将最短路径的形式与网络进行结合，同时通过四个通道融合了不同的异构信息。

d)BLSTM[29]。采用 NLP 工具和词法资源衍生出的众多特征与双向 LSTM 网络学习句子级别的特征。

e)Att-BLSTM[12]。提出了一种将注意力融合进双向循环神经网络的网络结构，提高了分类器对关键信息的注意力。

f)2ATT-BLSTM-BLSTM[22]。提出了一种融合注意力机制的叠层的双向循环神经网络模型。

g)BLSTM-BTLSTM[21]。提出了一种将实体识别与关系分类拼接在一起的端到端的神经网络模型，其中实体识别用 Bi-LSTM 来表示，关系分类用 Bi-Tree-LSTM 来表示。

如表 5 所示，相对于其他模型，在更少的额外的特征信息的情况下，H-RNN 的结果达到了目前最优效果。

4　结束语

本文针对现有基于特定句法结构的关系分类模型无法迁移到其他句法结构上的问题，提出了一种融合多句法结构的叠层循环神经网络模型。该模型通过利用 Bi-LSTM-CRF（融合 attention）和 Bi-Tree-LSTM 来表示文本序列和句法结构，将实体预训练和关系分类融合到端到端的框架中利用共享参数进行训练，在实体预训练阶段融合 attention 机制提升对实体的关注度，同时对多种句法结构进行加权学习。

在 Semeval-2010 Task8 数据集上实验证明，1)多句法结构的网络模型能够对不同句法结构的关系进行有效分类，具备一定的鲁棒性。同时，进一步证明了 SDP 结构(相对于其他结构)对于关系分类是最有效的。2)通过构造端到端的模型，在实体预训练阶段利用 attention 机制提升模型对实体的关注度，利用共享参数的学习方式，有效提升了模型的分类精度。

参考文献：

[1] Iris H, Su Nam K, Zornitsa K, *et al.* Semeval-2010 task 8: multi-way classification of semantic relations between pairs of nominals [C]// Proc of the 5th International Workshop on Semantic Evaluation, SemEval'10. Stroudsburg, PA: ACL Press, 2010: 33–38.

[2] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations [C]// Proc of the 42nd ACL on Interactive poster and demonstration sessions. Stroudsburg, PA: ACL Press, 2004 : No. 22.

[3] Zhou GuoDong, Su Jian, Zhang Jie, *et al.* Exploring various knowledge in relation extraction [C]// Proc of the 43rd annual meeting on association for computational linguistics. Stroudsburg, PA : ACL Press, 2005: 427-434.

[4] Zhang Zhu. Weakly-supervised relation classification for information extraction [C]// Proc of the 13th ACM international conference on Information and knowledge management. New York: ACM Press, 2004: 581-588.



[5] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction [J]. Journal of machine learning research, 2003, 3 (Feb): 1083-1106.

[6] Culotta A, Sorensen J. Dependency tree kernels for relation extraction [C]// Proc of the 42nd Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA : ACL Press, 2004: No. 423.

[7] Bunescu R C, Mooney R J. A shortest path dependency kernel for relation extraction [C]// Proc of Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL Press, 2005: 724-731.

[8] Katiyar A, Cardie C. Nested named entity recognition revisited [C]// Proc of the 16th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL Press, 2018: 861-871.

[9] Wang Qi, Xia Yuhang, Zhou Yangming, *et al.* Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition [J]. arXiv, 2018, preprint arXiv: 1804. 05017.

[10] Zheng Suncong, Wang Feng, Bao Hongyun, *et al.* Joint extraction of entities and relations based on a novel tagging scheme [C]// Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2017: 1227–1236.

[11] Yan Xu, Mou Lili, Li Ge, *et al.* Classifying relations via long short term memory networks along shortest dependency paths [C]// Proc of the 20th Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA : ACL Press, 2015: 1785-1794.

[12] Zhou Peng, Wei Shi, Tian Jun, *et al.* Attention-based bidirectional long short-term memory networks for relation classification [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA : ACL Press, 2016: 207-212.

[13] Zeng Daojian, Liu Kang, Lai Siwei, *et al.* Relation classification via convolutional deep neural network [C]// Proc of the 25th International Conference on Computational Linguistics. Stroudsburg, PA : ACL Press, 2014: 2335-2344.

[14] Aone C, Ramos-Santacruz M. REES: a large-scale relation and event extraction system [C]// Proc of the 6th conference on Applied natural language processing. Stroudsburg, PA : ACL Press, 2000 : 76-83.

[15] Zhang Min, Zhang Jie, Su Jian. Exploring syntactic features for relation extraction using a convolution tree kernel [C]// Proc of Main Conference on Human Language Technology, Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg, PA: ACL Press, 2006: 288-295.

[16] Zhou Guodong, Zhang Min, Ji Donghong, *et al.* Tree kernel-based relation extraction with context-sensitive structured parse tree information [C]// Proc of the Joint Conference on EMNLP-CoNLL. Stroudsburg, PA : ACL Press, 2007: 728-736.

[17] Socher R, Huval B, Manning C D, *et al.* Semantic compositionality through recursive matrix-vector spaces [C]// Proc of Joint Conference on EMNLP-CoNLL. Stroudsburg, PA : ACL Press, 2012: 1201-1211.

[18] Xu Kun, Feng Yansong, Huang Songfang, *et al.* Semantic relation classification via convolutional neural networks with simple negative sampling [J]. arXiv, 2015, preprint arXiv: 1506. 07650.

[19] Liu Yang, Wei Furu, Li Sujian, *et al.* A dependency-based neural network for relation classification [J]. arXiv, 2015, preprint arXiv: 1507. 04646.

[20] Li Jiwei, Thang Luong, Dan Jurafsky, *et al.* When are tree structures necessary for deep learning of representations? [C]// Proc of Conference on EMNLP. Stroudsburg, PA : ACL Press, 2015: 2304-2314.

[21] Miwa M, Bansal M. End-to-end relation extraction using lstms on sequences and tree structures [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA : ACL Press, 2016: 1105–1116.

[22] Xiao Minguang, Liu Cong. Semantic relation classification via hierarchical recurrent neural network with attention [C]// Proc of the 26th International Conference on Computational Linguistics. Stroudsburg, PA : ACL Press, 2016: 1254-1263.

[23] Zhang Xiaobin, Chen Fucai, Huang Ruiyang. A Combination of RNN and CNN for attention-based relation classification [J]. Procedia Computer Science, 2018, 131: 911-917.

[24] Feng Jun, Huang Minlie, Zhao Li, *et al.* Reinforcement Learning for Relation Classification from Noisy Data [C]// Proc of AAAI. Palo Alto, CA: AAAI Press, 2018.

[25] Liu Wenpeng, Cao Yanan, Cao Cong, *et al.* An adversarial training framework for relation classification [C]// Proc of ICCS. Berlin: Springer Press, 2018: 194-205.

[26] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks [C]// Proc of IJCNN on IEEE. Piscataway, NJ: IEEE Press, 2005, 4: 2047-2052.

[27] Ratinov L, Roth D. Design challenges and misconceptions in named entity recognition [C]// Proc of the 13th Conference on Computational Natural Language Learning. Stroudsburg, PA: ACL Press, 2009: 147-155.

[28] Rink B, Harabagiu S. Utd: Classifying semantic relations by combining lexical and semantic resources [C]// Proc of the 5th International Workshop on Semantic Evaluation. Stroudsburg, PA: ACL Press, 2010: 256-259.

[29] Zhang Shu, Zheng Dequan, Hu Xinchen, *et al.* Bidirectional long short-term memory networks for relation classification [C]// Proc of the 29th Pacific Asia Conference on Language, Information and Computation. Stroudsburg, PA : ACL Press, 2015: 73-78.